Data-driven Approaches for Social Video Distribution

Zhi Wang

Tsinghua University

June 29, 2015

Abstract

The Internet has recently witnessed the convergence of online social network services and online video services: users import videos from content sharing sites, and propagate them along the social connections by re-sharing them. Such social behaviors have dramatically reshaped how videos are disseminated, and the users are now actively engaged to be part of the social ecosystem, rather than being passively consumers. Despite the increasingly abundant bandwidth and computation resources, the ever increasing data volume of user generated video content and the boundless coverage of socialized sharing have presented unprecedented challenges. In this paper, we first presents the challenges in social-aware video delivery. Then, we present a principal framework for data-driven social video delivery approaches. Moreover, we identify the unique characteristics of social-aware video access and the social content propagation, and closely reveal the design of individual modules and their integration towards enhancing users' experience in the social network context.

1 Introduction

In the past decade, we have witnessed the fast evolution toward new generation networked multimedia processing and sharing in the Web 2.0 era. The users are now actively engaged to be part of a social ecosystem, rather than passively receiving media content. The revolution is being driven further by the deep penetration of 3G/4G wireless networks and smart mobile devices that are seamlessly integrated with online social networking and media sharing services.

Despite the increasingly abundant bandwidth and computation resources, the ever increasing data volume of user generated video content and the boundless coverage of socialized sharing have presented unprecedented challenges to both content and network service providers. The highly diversified content origins and distribution channels further complicate the design and management of online video sharing systems. This paper identifies the key issues and presenting our data-driven solution toward this promising research direction.
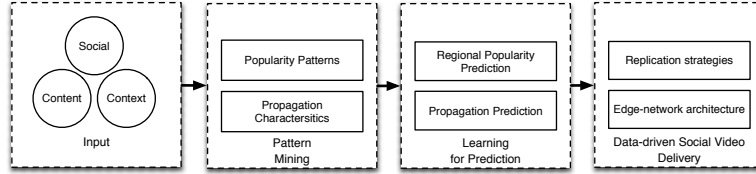
Figure 1: Framework of social-aware video content delivery.

1.1 Background and Challenges in Social Video Delivery

Online social network services connect users through "friendship" (*e.g.*, Facebook), "following" relationship (*e.g.*, Twitter), or professional connections (*e.g.*, LinkedIn). Such applications have successfully changed how people are connected to each other, and how they share information, including video. We have recently seen the convergence of online social network services and online video services: users can "import" videos from video sharing sites to online social networks, and make the videos propagate along the social connections by *re-sharing* them. The social behaviors have dramatically reshaped how videos are disseminated to users: *people are now receiving videos from friends directly.* For example, the online video clip "Gangnam Style", attracted over 1 billion views in 6 months after it is uploaded, due to its propagation over popular online social networks including Twitter and Facebook. Today, 500 years of video are watched and shared every day by Facebook users, and over 700 videos are shared on Twitter each minute [1].

Conventional video delivery strategies, *e.g.*, the original Client/Server streaming, IP and Application-Layer multicast / peer-to-peer, and Content Delivery Networks (CDNs), mainly focus on improving the network delivery performance, so as to meet the increasing scale of video requests. They have generally assumed that the content comes from centralized providers and the users only *passively* receive the contents [10]. For videos shared over social networks, however, the access patterns are much more dynamic, being affected by individuals and their activities during propagation. With information propagation, the scale/coverage in the social network context can be much larger and broader, making the system design much more involving.

More specifically, we are facing the following challenges in social video delivery.

First, users instead of content providers determine how videos reach each other. In an online social network, contents are generated, propagated, and disseminated by users. In 2014, YouTube reported that over 100 hours of video clips were uploaded by users in every minute. The content delivery systems thus have to distribute a much larger volume of user-generated videos than what has been handled by conventional content providers ever [3]. Second, the users share videos through social connections, and they tend to receive videos from their friends. As such, a provider no longer has tight and centralized control over the

dissemination of contents.

Second, dynamical content propagation. Social propagation is affected by a combination of factors, including the social topology, user behaviors, and the inherent content characteristics, to name but a few [9]. Given the many influential factors, propagation is highly dynamical, and traditional content delivery strategies generally lack prediction tools to infer such inherent dynamics.

Third, change of social content popularity. The change of the content origin and the social propagation also change the popularity distribution of videos. On one hand, the overall skewness of social video popularity distribution has been amplified; on the other hand, certain unpopular videos can be re-identified and then shared among users with close social relations. Video popularity is a key factor in designing and optimizing video delivery systems, and the change thus has strong implications. For example, it has been observed that the request hit ratio can be degraded by over 70% when traditional cache strategies were used to handle online social contents [5].

1.2 Data-driven Social Video Delivery Framework

These challenges demand a joint study on user behaviors, social video popularity, and social propagation to enhance the video content delivery in the social network context. To this end, Fig. 1 illustrates a general framework of social-aware video content delivery. This integrated framework takes the social, content and context information as the input to mine the popularity patterns and propagation characteristics. A machine learning model is designed to predict the evolution of social video popularity and the propagation patterns (*e.g.*, the size of a social cascade). The prediction reveals how videos would be shared at different regions, and the network resources can then be adaptively allocated among these regions. Content replication and caching strategies will be incorporated as well to further improve the sharing efficiency [6].

We next check the unique characteristics of social-aware video access and the social content propagation, and closely examine the design of individual modules and their integration in the framework.

2 Social Video Popularity

We start from the popularity of videos propagated through online social networks, including the distribution, evolution, and effective prediction.

2.1 Social-Aware Popularity Distribution

Since there is a large fraction of unpopular social videos, we further investigate in which types of social groups these unpopular videos propagate, using a Weibo (a Twitter-like social network) dataset that provides traces recording 400,000 videos shared by users in 1 month. By randomly sampling 50 videos
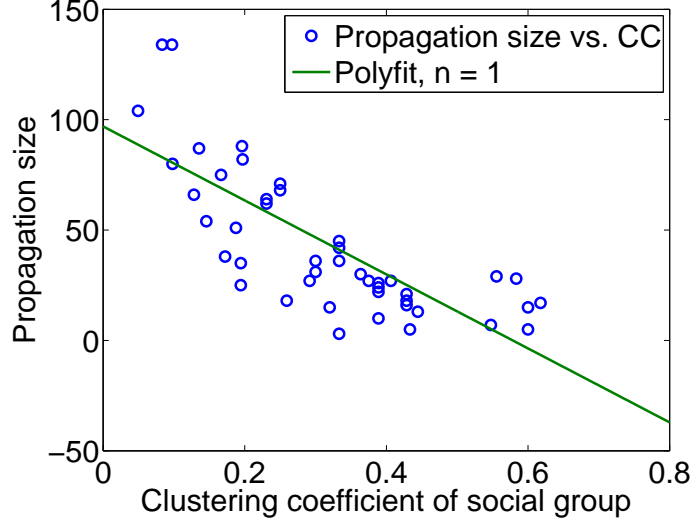
Figure 2: Unpopular videos tend to be shared in small social groups with membered socially connected.

with different propagation size (the number of users involved in a video's propagation), we explore the correlation between the propagation size and the clustering coefficient (larger clustering coefficient indicates close connection between members) of the social group formed by the users involved in the propagation. In Fig. 2, each sample illustrates the video propagation size versus the clustering coefficient of the corresponding social group. We observe a relatively strong correlation between the propagation size and the clustering coefficient; that is, unpopular videos tend to be shared among small social groups that are closely-connected (socially).

2.2 Social Popularity Prediction

The above observations suggest that content delivery mechanisms need to be substantially revised, and the social-related factors (*e.g.*, the total number of previous views and the video age) will play important roles [2]. In particular, the popularity prediction has to jointly consider both *accuracy* and *timeliness* of the prediction results.

As shown in [11], each video can be assigned with an index $k \in \{1, 2, ..., K\}$ according to the absolute time t_{init}^k when the video is initiated. Once a video is shared, it will be propagated through the social network for some time duration. A video has an age of $n \in \{1, 2, ...\}$ periods if it has been propagated through the social media for n periods. In each period, the video is further shared and viewed by users depending on the sharing and viewing status of the previous period. The propagation characteristics of video k up to age n are captured by

a d_n-dimensional vector $\mathbf{x}_n^k \in \mathcal{X}_n$ which includes information such as the total number of views and other contextual information, such as the characteristics of the social network over which the video propagates.

The popularity prediction can then be formulated as a multi-stage sequential decision and online learning problem [11]. Using multi-level popularity prediction in an online fashion, such a social-aware prediction outperforms existing view-based approaches by more than 30% in terms of prediction reward (a tradeoff between the popularity prediction accuracy and the timeliness).

3 Dynamic Social Video Propagation

The popularity reflects the macroscopic aggregated views of the videos shared. We next have a closer look at the social propagation process for videos, which determines how individual videos reach different users in online social networks.

3.1 Propagation Patterns

The generation and re-share of a social video typically form a propagation tree, rooted at the user who generates the video or initiates the sharing (referred to as the *initiator* or *root*). We refer to the users who re-share the video as *spreaders*, and the users who receive the shared video as *viewers* (or *receivers*). A video's popularity can then be calculated as the sum of its spreaders and receivers.

Large-scale measurement studies have also discovered interesting locality patterns in the propagation structures [8].

▷ *Geographical locality*: A large fraction of the videos are shared between users who are geographically close to each other. In Fig. 3, we plot the CDF of distances between users who join the social propagation of the same video in the online social network. Different curves are for videos with different popularities: (1) Popular videos: videos whose popularity is in the top 2%; (2) Unpopular videos: videos whose popularity is in the bottom 30%. We observe that different from traditional video consumption, unpopular social videos tend to be shared in local regions, where users are nearby to each other. For example around 40% of the distances between users sharing the same unpopular videos are close to 0km.

▷ *Temporal locality*: In the online social network, the users are more likely to re-share new video contents, *i.e.*, videos that are recently imported or re-shared. Fig. 4 illustrates the number of re-shares of a video in a timeslot (1 hour) versus the time lag since the propagation. We observe that most of the re-shares happen in the recent hours, and the re-share number against the time lag follows a zipf-like distribution with a shape parameter $s = 1.5070$. More than 95% of the re-shares happen within the first 24 hours. This indicates that in social video sharing, users' behaviors are highly crowded around the time point when it is imported.

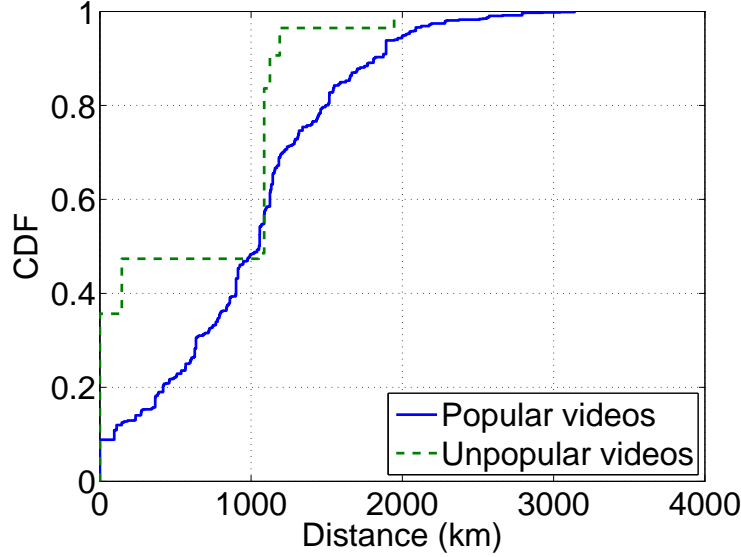Figure 3: CDF of distance between users that are sharing in the same propagation of social videos.

3.2 Propagation Model

Given the propagation patterns and the different types of nodes, an *SIR (Susceptible-Infectious-Recovered) model* can be used to model social video propagation. It considers a fixed population with three compartments: *susceptible*, *infectious*, and *recovered*. Li *et al.* [4] particularly investigate an extended epidemic model to capture the video propagation in online social networks:

Initially, a user shares this video from an external video sharing site and this initiator becomes *infectious*. All other users in the social network are *safe* except the friends of the initiator. The shared video appears in the news feed of the initiator's friends and thus they become *susceptible*. After a while, these friends log into the social network gradually and decide whether to watch the video (*infected*) or not (*immune*). For those *infected* users, they will decide whether to share after watching the video. They become *recovered* if they choose not to share. They will become Infectious if they choose to share. Again, these *infectious* users will make their friends who are on *safe* stage become *susceptible*.

Based on the propagation model, the parameters can be trained using real propagation traces, and the connection between video access and social activities is created for content delivery strategies.
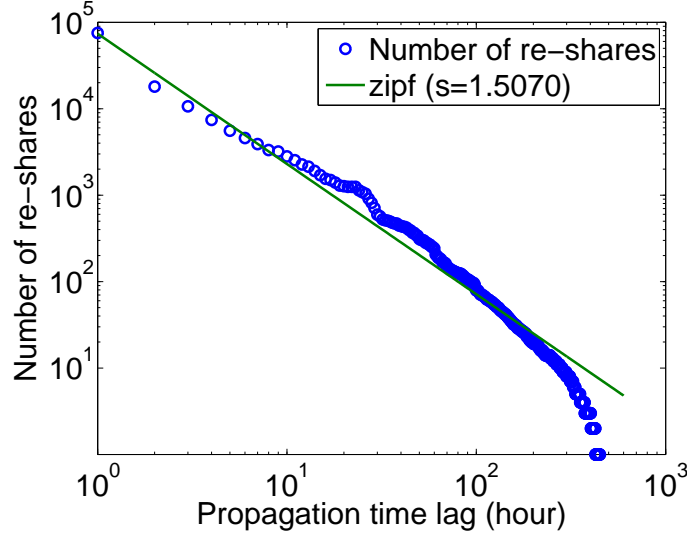
Figure 4: Number of re-shares versus the time lag.

4 Data-driven Social Video Delivery

Given the unique characteristics of the social videos, a series of solutions have been proposed in the literature, addressing the challenges in different aspects.

4.1 Propagation Prediction based Replication

A significant amount of efforts have been devoted to utilize social information to enhance the strategies in the traditional content delivery (*e.g.*, CDN, Peer-to-Peer Network). Wang *et al.* [8] proposed a hybrid CDN and P2P architecture for social video distribution, where the CDN servers can support the time-varying bandwidth and storage allocations requested by different regions, while peers are able to help contribute to each other in similar social groups.

In this replication architecture, two overlays exist in the social-aware replication design (as illustrated in Fig. 5): (1) *social propagation overlay* based on the social graph, which determines the video propagation among friends, *i.e.*, users after generating a video, can share the video with their direct friends, who will further re-share the video to more people, and (2) *delivery overlay* which determines how video contents are delivered from edge-cloud servers to users or among themselves in a peer-assisted paradigm.

Geographic Influence Index: The replication is designed to take social propagation into account. When performing video replication, we need to find out the videos that may propagate to more regions in the future. We design a geographic influence index in the region prediction for that, as below:

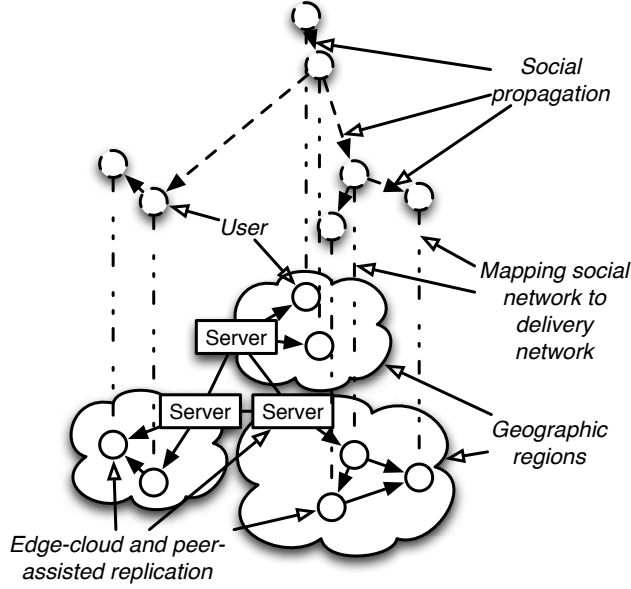$$g_v^{(T)} = c_1 \log(c_2 s_v^{(T-1)}),$$

Figure 5: Edge-cloud and peer-assisted social replication.

where $s_v^{(T-1)}$ is the propagation size of the propagation tree of video v in timeslot $T-1$. $g_v^{(T)}$ approximates the number of regions where the video will propagate to in the future time window, according to our measurement studies. Intuitively, a video should be replicated to more regions when the predicted number of regions involved in the propagation is larger than the number of regions it has already been replicated to (Θ). To proactively replicate social videos, a video with a geographic influence index $g_v^{(T)} > \Theta$ should be replicated to more regions to serve users locally. Parameters c_1 and c_2 are selected to fit the measurements. Based on the geographic influence index, we can predict whether the regions where the video has been replicated are enough.

4.2 Mobility and Propagation Prediction based Caching

Crowdsourced caching by replicating bandwidth-intensive videos on the *edge-network* devices (*e.g.*, users' smartphones) to let users serve each other is promising. Previous studies have demonstrated that such device-to-device (D2D) content sharing is possible when users are close to each other, and the contents to be delivered by users are delay tolerant [7]. However, in traditional D2D content sharing, a user broadcasts generated contents or reshared contents to a set of random users that are close to it. As a result, all contents are disseminated to users in the same way (*e.g.*, random flooding), causing the following problems: (1) In greedy flooding, smart devices in edge networks have to spend expansive power to cache and relay excessive contents. As the number of user-generated

8

(a) Social topology

(b) Content propagation

(c) User regional mobility

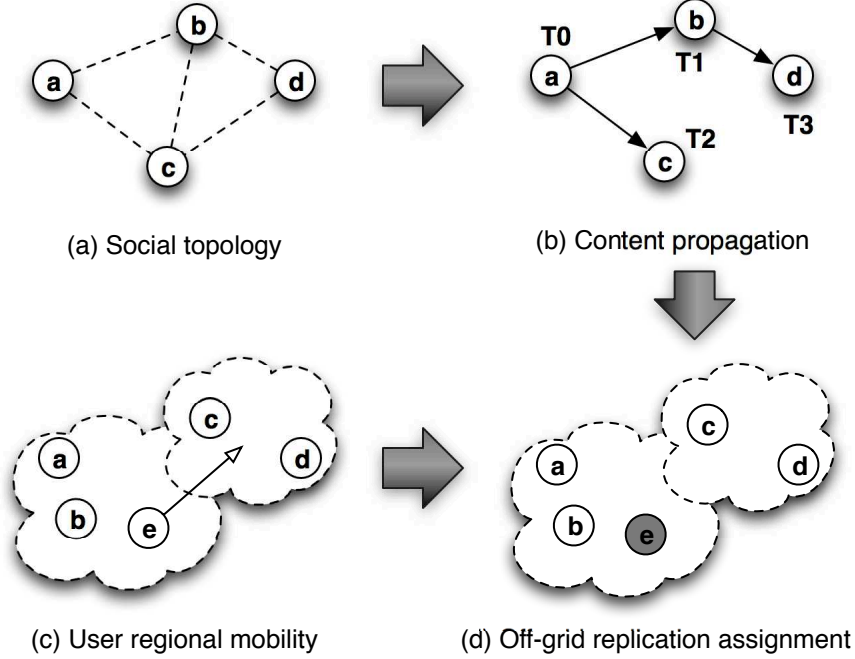(d) Off-grid replication assignment

Figure 6: Crowdsourced replication affected by social topology, content propagation and user regional mobility.

social contents is increasing, such mechanism is inherently non-scalable; (2) Social videos have heterogeneous popularity, while the conventional approaches treat them all the same, resulting in wasted resource to carry unpopular contents. (3) Due to the dynamic mobility patterns, users may not be able to fetch contents timely, resulting in poor quality of experience.

To address these issues, a joint propagation- and mobility-aware crowdsourced replication strategy can be developed based on social propagation characteristics and crowd mobility patterns in edge-network *regions*, *e.g.*, an area of hundred-meter range where users can move across. As illustrated in Fig. 6, using the social graph and propagation patterns, we first estimate how contents will be received by users, and we then predict to which regions the users will be moving and how long they will stay. Instead of letting contents flood between users that are merely close to each other, we disseminate social videos according to the influence of users and the propagation of videos. In this example, as user e – whilenot a friend of any other user – is moving to the region where user c and d are. Thus, e will be selected to replicate the content generated by user a, and both user c and d will receive the content shared by user a in the social propagation at times $T2$ and $T3$, respectively.

5 Summary and Future Direction

With the advances of online social networking, users, instead of content providers, determine how videos reach each other. Such key characteristics of networked videos as the popularity distribution and its evolution have been strongly affected by the social behaviors of the users, challenging the traditional content delivery that consider users as passive consumers only. In this paper, we identified the unique patterns and characteristics of social video propagation and content popularity through large-scale trace data analysis. We demonstrated a series of strategies of great potentials, including content replication, crowdsourced content caching, and network resource allocation. This new and promising research area opens many challenging issues to be addressed in the near future, and we believe that a data-driven framework design will be the key. Within this framework, deeply understanding on the user behaviors, the social propagation structures, the knowledge of content characteristics and context information, as well as social-relationship-based collaborative content sharing mechanisms will all play important roles.

References

[1] Youtube, http://www.youtube.com/yt/press/statistics.html.

[2] BORGHOL, Y., ARDON, S., CARLSSON, N., EAGER, D., AND MAHANTI, A. The Untold Story of the Clones: Content-agnostic Factors That Impact YouTube Video Popularity. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)* (2012).

[3] CHA, M., KWAK, H., RODRIGUEZ, P., AHN, Y., AND MOON, S. I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. In *ACM SIGCOMM* (2007).

[4] LI, H., CHENG, X., AND LIU, J. Understanding Video Sharing Propagation in Social Networks: Measurement and Analysis. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 10*, 4 (2014), 33.

[5] MISLOVE, A. Rethinking Web Content Distribution in the Social Media Era. In *NSF Workshop on Social Networks and Mobility in the Cloud* (2012).

[6] TORRES, R., FINAMORE, A., KIM, J. R., MELLIA, M., MUNAFO, M. M., AND RAO, S. Dissecting Video Server Selection Strategies in the YouTube CDN. In *IEEE International Conference on Distributed Computing Systems (ICDCS)* (2011).

[7] WANG, X., CHEN, M., HAN, Z., WU, D. O., AND KWON, T. T. Toss: Traffic offloading by social network service-based opportunistic sharing in

mobile social networks. In *IEEE International Conference on Computer Communications (INFOCOM)* (2014), IEEE, pp. 2346–2354.

[8] WANG, Z., SUN, L., CHEN, X., ZHU, W., LIU, J., CHEN, M., AND YANG, S. Propagation-based Social-aware Replication for Social Video Contents. In *ACM International Conference on Multimedia (Multimedia)* (2012), pp. 29–38.

[9] WANG, Z., SUN, L., WU, C., AND YANG, S. Enhancing Internet-scale Video Service Deployment Using Microblog-based Prediction. *IEEE Transactions on Parallel and Distributed Systems 26*, 3 (March 2015), 775–785.

[10] WANG, Z., SUN, L., ZHU, W., YANG, S., LI, H., AND WU, D. Joint Social and Content Recommendation for User Generated Videos in Online Social Network. *IEEE Transactions on Multimedia 15*, 3 (April 2013), 698–709.

[11] XU, J., VAN DER SCHAAR, M., LIU, J., AND LI, H. Forecasting Popularity of Videos using Social Media. *arXiv preprint arXiv:1403.5603* (2014).